IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant	:	Goddard, et al.
Appl. No.	:	10/063,566
Filed	:	May 2, 2002
For	:	SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC ACIDS ENCODING THE SAME
Examiner	:	Hunnicutt, Rachel Kapust
Group Art Unit	:	1647

DECLARATION UNDER 37 CFR §1.131

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

We declare and state as follows:

1. We are the inventors of the invention claimed in the above-captioned patent application.

2. During the time period in which all of the events and activities described herein occurred, we were employed by Genentech, Inc., the assignee of the above-captioned application.

3. All of the events and activities described herein were performed by us personally, or by others at our direction as part of our duties as employees of Genentech, Inc.

4. The invention claimed in the above-captioned patent application was conceived prior to April 29, 1998 and diligently reduced to practice thereafter in the U.S. as described below.

5. Prior to April 29, 1998, we conceived of the invention claimed in the above-captioned patent application. Prior to April 29, 1998, the idea of investigating several newly discovered DNA sequences for their relevance, including developing primers and cloning the DNA sequences of interest from normal and tumor tissues, was conceived. The attached sequence printout (Exhibit A), dated prior to April 29, 1998, shows the complete sequence of the nucleic acid having the sequence of SEQ ID NO:57, as well as the complete sequence of the amino acid of SEQ ID NO:58. Thus, conception of the invention claimed in the above-captioned patent application occurred prior to April 29, 1998.

6. The date deleted from Exhibit A is prior to April 29, 1998. This date was redacted pursuant to M.P.E.P. § 715.07. The date that remains is the date the report was printed, December 21, 2004.

7. After these initial experiments, we continued to produce primers, clone and sequence other DNA sequences. We then began to identify the expression levels of the cloned sequences, and created constructs for expression of the encoded proteins. The enclosed printed copy of an electronic file (Exhibit B) shows that PCR primers for numerous sequences which had been previously identified were designed on March 6, 2000 (as evidenced by the file name, "oli.out3_6_00"), including the primers for DNA59609, the DNA which is relevant to the presently claimed invention (see top of page 3 of Exhibit B). This data shows diligence in reducing to practice following conception of the invention. Thereafter, the primers were tested for expression in various normal and tumor tissues on June 13, 2000 (Exhibit C). These pages show a series of gels and a summary of results in which the primers were used to determine tissue and tumor expression levels for the various DNA sequences. The levels are shown using a ++, +, +-, and - to indicate the intensity of the specific signal which was detected. The data for DNA59609 is shown in row 10 on the results summary. This data shows that the DNA is more highly expressed in esophageal tumor as compared to normal esophagus tissue. Actual reduction to practice therefore occurred by at least June 13, 2000. Thus, we conceived of the present invention prior to April 29, 1998 and were diligent in reducing the invention to practice by at least June 13, 2000.

8. We hereby declare that all statements made herein of our own knowledge are true and that all statements made on information or belief are believed to be true, and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful statements may jeopardize the validity of the application or any patent issued thereon.

By: _____ Date: _____
Andrey Goddard

By: _____ Date: 2/15/05
Paul J. Godowski

By: _____ Date: _____
J. Christopher Grimaldi

By: _____ Date: _____
Austin L. Gurney

By: _____ Date: _____
William I. Wood

6. The date deleted from Exhibit A is prior to April 29, 1998. This date was redacted pursuant to M.P.E.P. § 715.07. The date that remains is the date the report was printed, December 21, 2004.

7. After these initial experiments, we continued to produce primers, clone and sequence other DNA sequences. We then began to identify the expression levels of the cloned sequences, and created constructs for expression of the encoded proteins. The enclosed printed copy of an electronic file (Exhibit B) shows that PCR primers for numerous sequences which had been previously identified were designed on March 6, 2000 (as evidenced by the file name, "oli.out3_6_00"), including the primers for DNA59609, the DNA which is relevant to the presently claimed invention (see top of page 3 of Exhibit B). This data shows diligence in reducing to practice following conception of the invention. Thereafter, the primers were tested for expression in various normal and tumor tissues on June 13, 2000 (Exhibit C). These pages show a series of gels and a summary of results in which the primers were used to determine tissue and tumor expression levels for the various DNA sequences. The levels are shown using a ++, +, +-, and - to indicate the intensity of the specific signal which was detected. The data for DNA59609 is shown in row 10 on the results summary. This data shows that the DNA is more highly expressed in esophageal tumor as compared to normal esophagus tissue. Actual reduction to practice therefore occurred by at least June 13, 2000. Thus, we conceived of the present invention prior to April 29, 1998 and were diligent in reducing the invention to practice by at least June 13, 2000.

8. We hereby declare that all statements made herein of our own knowledge are true and that all statements made on information or belief are believed to be true, and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful statements may jeopardize the validity of the application or any patent issued thereon.

By: _____ Date: Feb. 16/05
Audrey Goddard

By: _____ Date: _____
Paul J. Godowski

By: _____ Date: _____
J. Christopher Grimaldi

By: _____ Date: _____
Austin L. Gurney

By: _____ Date: _____
William I. Wood

6. The date deleted from Exhibit A is prior to April 29, 1998. This date was redacted pursuant to M.P.E.P. § 715.07. The date that remains is the date the report was printed, December 21, 2004.

7. After these initial experiments, we continued to produce primers, clone and sequence other DNA sequences. We then began to identify the expression levels of the cloned sequences, and created constructs for expression of the encoded proteins. The enclosed printed copy of an electronic file (Exhibit B) shows that PCR primers for numerous sequences which had been previously identified were designed on March 6, 2000 (as evidenced by the file name, "oli.out3_6_00"), including the primers for DNA59609, the DNA which is relevant to the presently claimed invention (see top of page 3 of Exhibit B). This data shows diligence in reducing to practice following conception of the invention. Thereafter, the primers were tested for expression in various normal and tumor tissues on June 13, 2000 (Exhibit C). These pages show a series of gels and a summary of results in which the primers were used to determine tissue and tumor expression levels for the various DNA sequences. The levels are shown using a ++, +, +-, and - to indicate the intensity of the specific signal which was detected. The data for DNA59609 is shown in row 10 on the results summary. This data shows that the DNA is more highly expressed in esophageal tumor as compared to normal esophagus tissue. Actual reduction to practice therefore occurred by at least June 13, 2000. Thus, we conceived of the present invention prior to April 29, 1998 and were diligent in reducing the invention to practice by at least June 13, 2000.

8. We hereby declare that all statements made herein of our own knowledge are true and that all statements made on information or belief are believed to be true, and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful statements may jeopardize the validity of the application or any patent issued thereon.

By: _____ Date: _____
Audrey Goddard

By: _____ Date: _____
Paul J. Godowski

By: _____ Date: 2/15/2005
J. Christopher Grimaldi

By: _____ Date: _____
Austin L. Gurney

By: _____ Date: _____
William I. Wood

6. The date deleted from Exhibit A is prior to April 29, 1998. This date was redacted pursuant to M.P.E.P. § 715.07. The date that remains is the date the report was printed, December 21, 2004.

7. After these initial experiments, we continued to produce primers, clone and sequence other DNA sequences. We then began to identify the expression levels of the cloned sequences, and created constructs for expression of the encoded proteins. The enclosed printed copy of an electronic file (Exhibit B) shows that PCR primers for numerous sequences which had been previously identified were designed on March 6, 2000 (as evidenced by the file name, "oli.out3_6_00"), including the primers for DNA59609, the DNA which is relevant to the presently claimed invention (see top of page 3 of Exhibit B). This data shows diligence in reducing to practice following conception of the invention. Thereafter, the primers were tested for expression in various normal and tumor tissues on June 13, 2000 (Exhibit C). These pages show a series of gels and a summary of results in which the primers were used to determine tissue and tumor expression levels for the various DNA sequences. The levels are shown using a ++, +, +-, and - to indicate the intensity of the specific signal which was detected. The data for DNA59609 is shown in row 10 on the results summary. This data shows that the DNA is more highly expressed in esophageal tumor as compared to normal esophagus tissue. Actual reduction to practice therefore occurred by at least June 13, 2000. Thus, we conceived of the present invention prior to April 29, 1998 and were diligent in reducing the invention to practice by at least June 13, 2000.

8. We hereby declare that all statements made herein of our own knowledge are true and that all statements made on information or belief are believed to be true, and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful statements may jeopardize the validity of the application or any patent issued thereon.

By: _____ Date: _____
Audrey Goddard

By: _____ Date: _____
Paul J. Godowski

By: _____ Date: _____
J. Christopher Grimaldi

By: _____ Date: 2/5/05
Austin L. Gurney

By: _____ Date: _____
William I. Wood

6. The date deleted from Exhibit A is prior to April 29, 1998. This date was redacted pursuant to M.P.E.P. § 715.07. The date that remains is the date the report was printed, December 21, 2004.

7. After these initial experiments, we continued to produce primers, clone and sequence other DNA sequences. We then began to identify the expression levels of the cloned sequences, and created constructs for expression of the encoded proteins. The enclosed printed copy of an electronic file (Exhibit B) shows that PCR primers for numerous sequences which had been previously identified were designed on March 6, 2000 (as evidenced by the file name, "oli.out3_6_00"), including the primers for DNA59609, the DNA which is relevant to the presently claimed invention (see top of page 3 of Exhibit B). This data shows diligence in reducing to practice following conception of the invention. Thereafter, the primers were tested for expression in various normal and tumor tissues on June 13, 2000 (Exhibit C). These pages show a series of gels and a summary of results in which the primers were used to determine tissue and tumor expression levels for the various DNA sequences. The levels are shown using a ++, +, +-, and - to indicate the intensity of the specific signal which was detected. The data for DNA59609 is shown in row 10 on the results summary. This data shows that the DNA is more highly expressed in esophageal tumor as compared to normal esophagus tissue. Actual reduction to practice therefore occurred by at least June 13, 2000. Thus, we conceived of the present invention prior to April 29, 1998 and were diligent in reducing the invention to practice by at least June 13, 2000.

8. We hereby declare that all statements made herein of our own knowledge are true and that all statements made on information or belief are believed to be true, and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful statements may jeopardize the validity of the application or any patent issued thereon.

By: _____ Date: _____
Audrey Goddard

By: _____ Date: _____
Paul J. Godowski

By: _____ Date: _____
J. Christopher Grimaldi

By: _____ Date: _____
Austin L. Gurney

By: William I Wood Date: 2/14/05
William I. Wood

EXHIBIT A

>Tuesday, December 21, 2004
>DNA59609 [Full]
>1380 Sites [All Sites]

DNA59609 godderda GSeqEdit

>Sequenced by transposon tagging. ACGT Peter Ma and Ellson Chen.

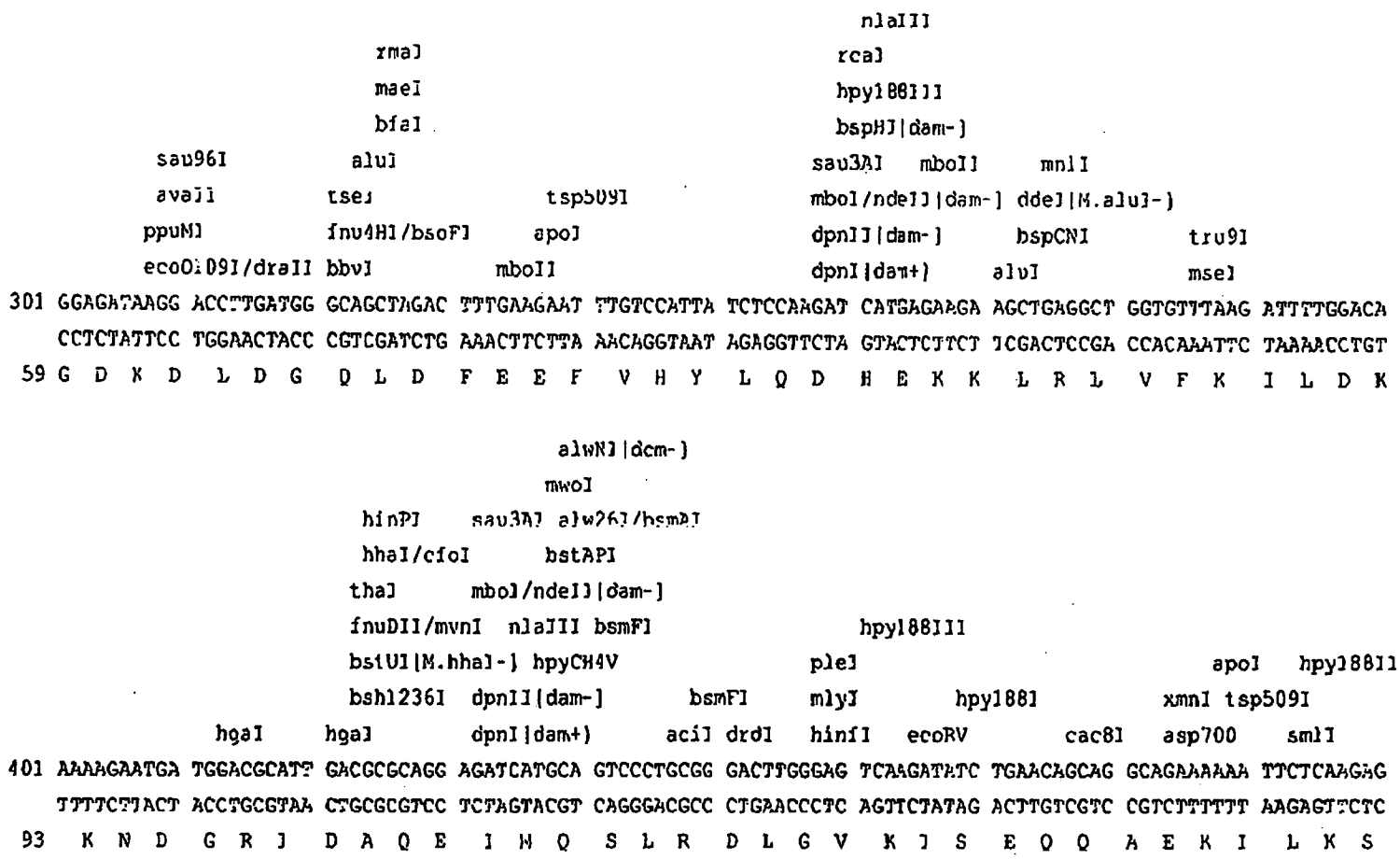
```
                                                                                   taqI
                                                                                 sstI
                                                                                 sacI
                                                                                 hgiAI/aspHI|M.aluI-]
                                                                                 ecl136II
                                                                                 bsp1286|M.aluI-]
                                                                                 bsiHKAI
                                             thaI                              xhoI   xhoI
                                        nlaIII   rsaI                          tliI   tliI
                                       sphI     fnuDII/mvnI               taqI  smlI   smlI
                                       nspHI    bstUI taiI               xhoI mwoI   bmyI              taqI
                                   taiI nspI    bsh1236I                 tliI fnu4HI/bsoFI    mwoI       xhoI
                                   maeII/hpyCH4IV bsiWI/splI        tsp509I|M.ecoRI-]  banII|M.aluI-]    tliI
                             aluI  hinlI/acyI cacBI csp6I          ecoRI  smlI aciI taqI paeR7I  tfiI    smlI
                     sfcI  sapI    ahaII/bsaHI  mluI  snaBI   hpy188I   paeR7I  paeR7I avaI|M.taqI-]   paeR7I  ts
                tsp45I     mboII   aatII cacBI  aflIII maeII/hpyCH4IV   avaI|M.taqI-] aluI   bglI     avaI|M.taqI
                maeIII     earI/ksp632I hpy99I hpyCH4V bsaAI aluI  apoI  mwoI bsrBI avaI|M.taqI-]  hinfI mwoI mnlI bts
              1 ATGTGACACT ATAGAAGAGC TATGACGTCG CATGCACGCG TACGTAAGCT CGGAATTCGG CTCGAGCGGC TCGAGCTCGA GCCGAATCGG CTCGAGGGGC
                TACACTGTGA TATCTTCTCG ATACTGCAGC GTACGTGCGC ATGCATTCGA GCCTTAAGCC GAGCTCGCCG AGCTCGAGCT CGGCTTAGCC GAGCTCCCCG
                                                                              ^start insert
```

>Tuesday, December 21, 2004
>DNA59609 |Full]
>1380 Sites |All Sites]
>Sequenced by transposon tagging. ACGT Peter Ma and Ellson Chen.

```
                                                                                    taqI
                                                                                sstI
                                                                                sacI
                                                                                hgiAI/aspHI|M.aluI-]
                                                                                ecl136II
                                                                                bsp1286|M.aluI-]
                                                                                bsiHKAI
                                         thaI                               xhoI   xhoI
                                   nlaIII   rsaI                            tliI   tliI
                                 sphI    fnuDII/mvnI                taqI    smlI   smlI
                                 nspHI   bstUI taiI                xhoI mwoI    bmyI                 taqI
                            taiI nspI    bsh1236I                  tliI fnu4HI/bsoFI     mwoI       xhoI
                            maeII/hpyCH4IV bsiWI/splI      tsp509I|M.ecoRI-]    banII|M.aluI-]      tliI
                    aluI   hinlI/acyI cac8I csp6I         ecoRI    smlI aciI  taqI paeR7I   tfiI    smlI
          sfcI  sapI       ahaII/bsaHI  mluI  snaBI   hpy188I      paeR7I   paeR7I avaI|M.taqI-]    paeR7I   ts
    tsp45I      mboII      aatII cac8I  aflIII maeII/hpyCH4IV      avaI|M.taqI-] aluI    bglI       avaI|M.taqI
    maeIII      earI/xsp632I hpy99I  hpyCH4V  bsaAI aluI  apoI   mwoI bsrBI avaI|M.taqI-]   hinfI mwoI mnlI bts
1 ATGTGACACT ATAGAAGAGC TATGACGTCG CATGCACGCG TACGTAAGCT CGGAATTCGG CTCGAGCGGC TCGAGCTCGA GCCGAATCGG CTCGAGGGGC
  TACACTGTGA TATCTTCTCG ATACTGCAGC GTACGTGCGC ATGCATTCGA GCCTTAAGCC GAGCTCGCCG AGCTCGAGCT CGGCTTAGCC GAGCTCCCCG
                                                                         ^start insert
```

```
                         nlaIII                        taiI  mspI                                         nl
    hgiAI/aspHI   aciI  nspHI                        maeII/hpyCH4IV                           rsaI      taqI  bsp
    bsp1286      fnu4HI/bsoFI                       bsaAI  hpaII                              csp6I  pleI     bmy
    bsiHKAI     haeIII/palI                        rsaI   bslI                                scaI   mlyI     ban
    bmyI      cac8I      nspI                      csp6I cfr10I/bsrFI                    bsrI        hinfI|M.taqI-
101 AGTGGAGCAC CCAGCAGGCC GCCAACATGC TCTGTCTGTG CCTGTACGTG CCGGTCATCG GGGAAGCCCA GACCGAGTTC CAGTACTTTG AGTCGAAGGG
    TCACCTCGTG GGTCGTCCGG CGGTTGTACG AGACAGACAC GGACATGCAC GGCCAGTAGC CCCTTCGGGT CTGGCTCAAG GTCATGAAAC TCAGCTTCCC
1                            M  L   C  L  C   L  Y  V   P  V  I  G   E  A  Q   T  E  F   Q  Y  F  E   S  K  G
                             ^Met

                                                              ecoRI
                                                            scrFI|dcm-)
                                                            pspGI
                                                            mvaI
                                                            ecoRII|dcm-)
                                                            dsaV|dcm-)
                                              bpuAI         bstNI
                                              bbsI  fokI    bssKI|dcm-)                                  rsaI
                                          ddeI|M.aluI-)  mnlI apyI|dcm+)                                 csp6I  bpm
              eco57I                      bspCNI mboII  bslI|dcm-) apoI         aciI tspRI               bsp1407I/bs
             aluI                      aluI tspRI    bstF5I   bsaJI tsp509I|M.ecoRI-) bsrI           tsp509I   aluI
201 GCTCCCTGCC GAGCTGAAGT CCATTTTCAA GCTCAGTGTC TTCATCCCCT CCCAGGAATT CTCCACCTAC CGCCAGTGGA AGCAGAAAAT TGTACAAGCT
    CGAGGGACGG CTCGACTTCA GGTAAAAGTT CGAGTCACAG AAGTAGGGGA GGGTCCTTAA GAGGTGGATG GCGGTCACCT TCGTCTTTTA ACATGTTCGA
26   L  P  A   E  L  K  S   I  F  K   L  S  V   F  I  P  S   Q  E  F   S  T  Y   R  Q  W  K   Q  K  I   V  Q  A
```


nlaIII
rmaI
rcaI
maeI
hpy188III
bfaI
bspHI|dam-)
sau96I
aluI
sau3AI mboII mnlI
avaII
tseI
tsp509I
mboI/ndeII|dam-] ddeI|M.aluI-)
ppuMI
fnu4HI/bsoFI
apoI
dpnII|dam-] bspCNI tru9I
eco0109I/draII bbvI
mboII
dpnI|dam+) aluI mseI
301 GGAGATAAGG ACCTTGATGG GCAGCTAGAC TTTGAAGAAT TTGTCCATTA TCTCCAAGAT CATGAGAAGA AGCTGAGGCT GGTGTTTAAG ATTTTGGACA
CCTCTATTCC TGGAACTACC CGTCGATCTG AAACTTCTTA AACAGGTAAT AGAGGTTCTA GTACTCTTCT TCGACTCCGA CCACAAATTC TAAAACCTGT
59 G D K D L D G Q L D F E E F V H Y L Q D H E K K L R L V F K I L D K
alwNI|dcm-)
mwoI
hinPI sau3AI alw26I/bsmAI
hhaI/cfoI bstAPI
thaI mboI/ndeII|dam-]
fnuDII/mvnI nlaIII bsmFI hpy188III
bstUI|M.hhaI-) hpyCH4V pleI apoI hpy188II
bshI236I dpnII|dam-] bsmFI mlyI hpy188I xmnI tsp509I
hgaI hgaI dpnI|dam+) aciI drdI hinfI ecoRV cac8I asp700 smlI
401 AAAAGAATGA TGGACGCATT GACGCGCAGG AGATCATGCA GTCCCTGCGG GACTTGGGAG TCAAGATATC TGAACAGCAG GCAGAAAAAA TTCTCAAGAG
TTTTCTTACT ACCTGCGTAA CTGCGCGTCC TCTAGTACGT CAGGGACGCC CTGAACCCTC AGTTCTATAG ACTTGTCGTC CGTCTTTTTT AAGAGTTCTC
93 K N D G R I D A Q E I M Q S L R D L G V K I S E Q Q A E K I L K S

```
                                                                                                  fokI
                                                                                 dsaI             bstF5I
                                                                                 btgI/bstDSI      sau3AI mnlI
                            taqI                                       mnlI draIII          avaI  mboI/ndeII|dam-|
                       mamI  pflMI                                   bseRI pflNI        fokI      dpnII|dam-]
     nlaIII      bceAI bsabI bslI  bsrI          bsmAI             mnlI  bslI bsaJI     bstF5I    dpnI|dam+|  bsrI
501  CATGGATAAA AACGGCACGA TGACCATCGA CTGGAACGAG TGGAGAGACT ACCACCTCCT CCACCCCGTG GAAAACATCC CCGAGATCAT CCTCTACTGG
     GTACCTATTT TTGCCGTGCT ACTGGTAGCT GACCTTGCTC ACCTCTCTGA TGGTGGAGGA GGTGGGGCAC CTTTTGTAGG GGCTCTAGTA GGAGATGACC
126  M  D  K   N  G  T  M   T  I  D   W  N  E   W  R  D  Y   H  L  L   H  P  V   E  N  I  P   E  I  I   L  Y  W

                                               mspI
                                               hpaII                                                     scrFI|dcm
                                               scrFI|M.hpaII-]                                           pspGI
                                               nciI                                                      mvaI
                                               dsaV                                                      ecoRII|dc
                                               bssKI                                                     dsaV|dcm-
                                             bsmFI                                                       bstNI
                 sau3AI                      sau96I                                                      bssKI|dcm
                 mboI/ndeII|dam-]            nlaIV  fokI               mnlI mnlI                         apyI|dcm+
                 dpnII|dam-]          tfiI   avaII  bstF5I       tspRI bseRI            fokI           draIII
      bsmI       dpnI|dam+]      hphI hinfI  bst4CI/hpyCH4III    bst4CI/hpyCH4III       bstF5I    bsmAI sexAI
601  AAGCATTCCA CGATCTTTGA TGTGGGTGAG AATCTAACGG TCCCGGATGA GTTCACAGTG GAGGAGAGGC AGACGGGGAT GTGGTGGAGA CACCTGGTGG
     TTCGTAAGGT GCTAGAAACT ACACCCACTC TTAGATTGCC AGGGCCTACT CAAGTGTCAC CTCCTCTCCG TCTGCCCCTA CACCACCTCT GTGGACCACC
159  K  H  S  T   I  F  D   V  G  E   N  L  T  V   P  D  E   F  T  V   E  E  R  Q   T  G  M   W  W  R   H  L  V  A
```

```
                                                  scrFI|dcm-]
                                                  pspGI
                                                  mvaI
                                                  ecoRII|dcm-}
                                                  dsaV|dcm-]
                                                  bstNI
                                  alwNI|dcm-]     bssKI[dcm-]
                                  alw26I/bsmAI sau96I|M.haeIII-]
                        bceAI  hpy188III      nlaIV  apyI|dcm+)      hgiAI/aspHI  sau96I        tseI
                       haeIII/paII       hpyCH4V    bslI             bsp1286  bspMI             fnu4HI/bsoFI
                       sau96I|M.haeIII-} bsgI  haeIII/paII      smlI  bsiHKAI hpyCH4V  nlaIII     bbvI          sf
      mnlI        nlaIV  bciVI    bspMI  bceAI   bsaJI      mwoI   bmyI  nlaIII avaII   mnlI acil        nlaIII
701 CAGGAGGTGG GGCAGGGGCC GTATCCAGAA CCTGCACGGC CCCCCTGGAC AGGCTCAAGG TGCTCATGCA GGTCCATGCC TCCCGCAGCA ACAACATGGG
    GTCCTCCACC CCGTCCCCGG CATAGGTCTT GGACGTGCCG GGGGGACCTG TCCGAGTTCC ACGAGTACGT CCAGGTACGG AGGGCGTCGT TGTTGTACCC
193   G  G  G   A  G  A   V  S  R  T   C  T  A   P  L  D   R  L  K  V   I  M  Q   W  H  A   S  R  S  N   N  M  G
```

```
                                           tsp45I
                                         scrFI|dcm-)
                                         pspGI
                                         mvaI
                                         ecoRII|dcm-)
                                         dsaV|dcm-)
                                         bstNI
                                         bssKI|dcm-)
                 hpy188I   taqI        haeIII/palI
                 ddeI      hpy188III  nlaIV   maeIII                      mnlI
                 bspCNI  tfiI       mnlI   apyI|dcm+)   aciI bsrDI       taiI     tsp509I    tfiI
               mslI      hinfI|M.taqI-) sau96I|dcm-)|M.haeIII-)    sfaNI  maeII/hpyCH4IV      hinfI  xcm
801 CATCGTTGGT GGCTTCACTC AGATGATTCG AGAAGGAGGG GCCAGGTCAC TCTGGCGGGG CAATGGCATC AACGTCCTCA AAATTGCCCC CGAATCAGCC
    GTAGCAACCA CCGAAGTGAG TCTACTAAGC TCTTCCTCCC CGGTCCAGTG AGACCGCCCC GTTACCGTAG TTGCAGGAGT TTTAACGGGG GCTTAGTCGG
226  I  V  G   G  F  T  Q   M  I  R   E  G  G   A  R  S  L   W  R  G   N  G  I   N  V  L  K   I  A  P   E  S  A

                                                   scrFI|dcm-)
                                                   pspGI    hpy188I
                                                   mvaI   pleI                               styI
                         hinPI                     ecoRII|dcm-)                              bslI
                         hhaI/cfoI                 dsaV|dcm-)                                bsaJI
                     sau3AI                        bstNI  mlyI                             sau96I
                     mboI/ndeII|dam-)              bssKI|dcm-)    tfiI                     avaII
                     dpnII|dam-)                   apyI|dcm+) ddeI  hinfI                  ppuMI         sau9
     tsp509I haeIII/palI dpnI|dam+)              tsp45I    hinfI mnlI  bssSI mnlI          nlaIV         nlaIV
    apoI nlaIII        mwoI    haeII             maeIII  bsmAI bspCNI  hpy188III           ecoO109I/draII hae
901 ATCAAATTCA TGGCCTATGA GCAGATCAAG CGCCTTGTTG GTAGTGACCA GGAGACTCTG AGGATTCACG AGAGGCTTGT GGCAGGGTCC TTGGCAGGGG
    TAGTTTAAGT ACCGGATACT CGTCTAGTTC GCGGAACAAC CATCACTGGT CCTCTGAGAC TCCTAAGTGC TCTCCGAACA CCGTCCCAGG AACCGTCCCC
259 I  K  F  M   A  Y  E   Q  I  K   R  L  V  G   S  D  Q   E  T  L   R  I  H  E   R  L  V   A  G  S   L  A  G  A
```

```
                                                                                                    xcmI  nl
                                                                                                    scrFI|dc
                                                                                                    pspGI bs
                                               fokI                                                 mvaI  ba
                                        mboII  mspI      tseI                                       ecoRII|d
                                      eco57I   hpaII     fnu4HI/bsoFI          hpy188III            dsaV|dcm
                                   sau96I    scrFI|M.hpaII-|  mboII         rsaI                    bstNI al
                 sfaNI             avaII bpuAI nciI     hinPI   bpuAI       csp6I                   bssKI|dc
              tseI                ppuMI  bbsI  dsaV     hhaI/cfoI         scaI ddeI                 apyI|dcm
              fnu4HI/bsoFI        eco0109I/draII  bstF5I  bbvI  bbsI    bsrI   bspCNI            hinPI  mnlI
              bbvI       bslI   mnlI hpy188III bssKI  haeII  aciI     haeIII/palI   bsmI         hhaI/cfoI
1001 CCATCGCCCA GAGCAGCATC TACCCAATGG AGGTCCTGAA GACCCGGATG GCGCTGCGGA AGACAGGCCA GTACTCAGGA ATGCTGGACT GCGCCAGGAG
     GGTAGCGGGT CTCGTCGTAG ATGGGTTACC TCCAGGACTT CTGGGCCTAC CGCGACGCCT TCTGTCCGGT CATGAGTCCT TACGACCTGA CGCGGTCCTC
293   I  A  Q   S  S  I   Y  P  M  E   V  L  K   T  R  M   A  L  R  K   T  G  Q   Y  S  G   M  L  D  C   A  R  R
```

```
        haeIII/pal]
       mscI/bal]|dcm-]
       eaeI[dcm-]
       cfrI
      scrFI|dcm-)
      pspGI
      mval xcmI
      ecoRII|dcm-]
      dsaV|dcm-]
      bstNI
      bssKI|dcm-]
      apyI|dcm+)                                                    sfaNI
    sau3AI  bslI          aciI                                    mspI
    mboI/ndeII|dam-)      fnu4HI/bsoFI          nlaIII            hpaII   drdI          hg
    dpnII|dam-]           haeIII/pall           nspHI      fokI   nae]/ngoMI            esp3
    dpnI|dam+]            eaeI                  bstXI      bstF5I cfr10I/bsrFI    accI  bsmB
    alwI|dam-]   mnlI     cfrI                  bsmFI nspI sfaNI  bslI cac8I  taqI hpyCH4V bsmA
1101 GATCCTGGCC AGAGAGGGGG TGGCCGCCTT CTACAAAGGC TATGTCCCCA ACATGCTGGG CATCATCCCC TATGCCGGCA TCGACCTTGC AGTCTACGAG
     CTAGGACCGG TCTCTCCCCC ACCGGCGGAA GATGTTTCCG ATACAGGGGT TGTACGACCC GTAGTAGGGG ATACGGCCGT AGCTGGAACG TCAGATGCTC
 326 I  L  A  R  E  G  V  A  A  F  Y  K  G  Y  V  P  N  M  L  G  I  I  P  Y  A  G  I  D  L  A  V  Y  E
```

```
                    tseI                        mspI
                    fnu4HI/bsoFI                hpaII
                  sfcI                          scrFI|M.hpaII-]
                 tseI                           nciI              scrFI|dcm-]
                 fnu4HI/bsoFI                 sau96I              pspGI
                 bbvI                         nlaIV               mvaI
               scrFI|dcm-]                    avaII               ecoRII|dcm-]
               pspGI   bbvI                  aciI  dsaV           dsaV|dcm-]                          pv
               mvaI pstI|M.HI-]             thaI   bssKI          bstNI                              haeII
               ecoRII|dcm-]                 fnuDII/mvnI           bssKI|dcm-]          rsaI    mscI/b
               dsaV|dcm-]                   bstUI|M.hhaI-]      hgiAI/aspHI           alwNI|dcm-) cac8
               bstNI hpyCH4V     tspRI      bsh1236I            bsp1286       nlaIV   alw26I/bsmAI  ms
        hpy188III bssKI|dcm-]      btsI     hinPI  bsaJI        bsiHKAI haeIII/palI      bsrI      eaeI
      sml1 bsmI  apyI|dcm+]     hpyCH4V     hhaI/cfoI           bmyI apyI|dcm+] banI  nlaIII  csp6I   cfrI
1201 ACGCTCAAGA ATGCCTGCCT GCAGCACTAT GCAGTGAACA GCGCGGACCC CGCCGTGTTT GTGCTCCTGC CCTGTGGCAC CATCTCCAGT ACCTGTGCGC
     TGCGAGTTCT TACGGACCGA CGTCGTGATA CGTCACTTGT CGCGCCTGGG GCCGCACAAA CACGAGGACC GGACACCGTG GTACAGGTCA TGGACACCGG
 359 T  L  K  N   A  W  L   Q  H  Y   A  V  N  S   A  D  P   G  V  F   V  L  L  A   C  G  T   M  S  S   T  C  G  Q
```

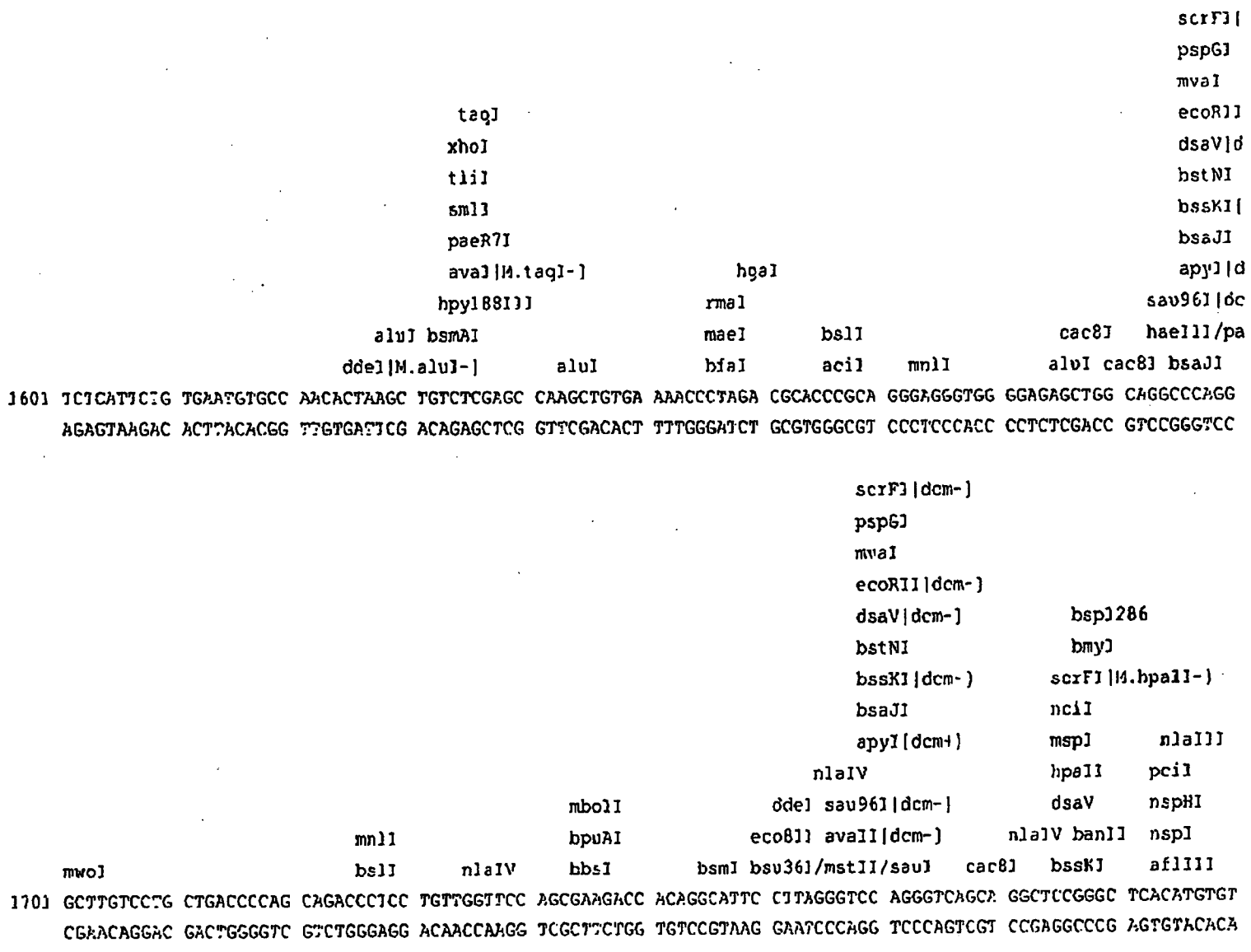
```
                                                 mspI
                                                 hpaII
                                   sau96I|dcm-]|M.haeIII-]
                                   haeIII/paII     scrFI|M.hpaII-]                         mspI
                                 scrFI|dcm-]       nciI                                   mroI
                                 pspGI   rmaI      dsaV                                   hpy188III
                         xcmI    mvaI    maeI      bssKI                                  bspMII
                       haeIII/paII  ecoRII|dcm-]  sau96I                                  bspEI
                      mscI/balI    dsaV|dcm-]    nlaIV  sfaNI                             bsaWI  tsp45I
                      eaeI         bstNI   bfaI  avaII fokI cac8I                         accIII maeIII      tseI   mboII
                      cfrI         bssKI|dcm-]  ppuMI bslI hpyCH4V cac8I              hinPI    mnlI      fnu4HI/bsoFI
                    cac8I  aluI    apyI|dcm+]   eco0109I/draII  hinPI  mnlI          haeII hpaII hphI  nlaIII   mnlI
                    aluI cac8I   bsaJI  bslI  hpy188III bstF5I  hhaI/cfoI        mnlI hhaI/cfoI  bstEII    bbvI earI/ksp632I
1301 AGCTGGCCAG CTACCCCCTG GCCCTAGTCA GGACCCGGAT GCAGGCGCAA GCCTCTATTG AGGGCGCTCC GGAGGTGACC ATGAGCAGCC TCTTCAAACA
     TCGACCGGTC GATGGGGGAC CGGGATCAGT CCTGGGCCTA CGTCCGCGTT CGGAGATAAC TCCCGCGAGG CCTCCACTGG TACTCGTCGG AGAAGTTTGT
393   L  A  S   Y  P  L   A  L  V  R   T  R  M   Q  A  Q   A  S  I  E   G  A  P   E  V  T   M  S  S  L   F  K  H

                          sau96I   nlaIV                                 nlaIII
                          avaII    eco0109I/draII  rsaI        sau96I|M.haeIII-]        aluI              taiI
                        rsrII/cspI                 csp6I       nlaIV        rcaI        pvuII             maeII/hpyCH4IV
                        cpoI    mnlI haeIII/paII bsp1407I/bsrGI haeIII/paII hpy188III fokI mspAII/nspBII    bsaAI
                       aciI  bsaJI sau96I|M.haeIII-]       cac8I        bspHI     bstF5I        sfaNI aluI      accI
1401 TATCCTGCGG ACCGAGGGGG CCTTCGGGCT GTACAGGGGG CTGGCCCCCA ACTTCATGAA GGTCATCCCA GCTGTGAGCA TCAGCTACGT GGTCTACGAG
     ATAGGACGCC TGGCTCCCCC GGAAGCCCGA CATGTCCCCC GACCGGGGGT TGAAGTACTT CCAGTAGGGT CGACACTCGT AGTCGATGCA CCAGATGCTC
426  I  L  R   T  E  G  A   F  G  L   Y  R  G   L  A  P  N   F  M  K   V  I  P   A  V  S  I   S  Y  V   V  Y  E
```

1501 AACCTGAAGA TCACCCTGGG CGTGCAGTCG CGGTGACGGG GGGAGGGCCG CCCGGCAGTG GACTCGCTGA TCCTGGGCCG CAGCCTGGGG TGTGCAGCCA

TTGGACTTCT AGTGGGACCC GCACGTCAGC GCCACTGCCC CCCTCCCGGC GGGCCGTCAC CTGAGCGACT AGGACCCGGC GTCGGACCCC ACACGTCGGT

459 N L K I T L G V Q S R Q


scrFI|
pspGI
mvaI
taqI
ecoRII
xhoI
dsaV|d
tliI
bstNI
smlI
bssKI[
paeR7I
bsaJI
aval|M.taql-]
hgaI
apyI|d
hpy188III
rmaI
sau96I|dc
aluI bsmAI
maeI
bsII
cac8I
haeIII/pa
ddel|M.aluI-]
aluI
bfaI
aciI
mnlI
alwI cac8I bsaJI
1601 TCTCATTCTG TGAATGTGCC AACACTAAGC TGTCTCGAGC CAAGCTGTGA AAACCCTAGA CGCACCCGCA GGGAGGGTGG GGAGAGCTGG CAGGCCCAGG
AGAGTAAGAC ACTTACACGG TTGTGATTCG ACAGAGCTCG GTTCGACACT TTTGGGATCT GCGTGGGCGT CCCTCCCACC CCTCTCGACC GTCCGGGTCC
scrFI|dcm-]
pspGI
mvaI
ecoRII|dcm-]
dsaV|dcm-]
bspI286
bstNI
bmyI
bssKI|dcm-)
scrFI|M.hpaII-)
bsaJI
nciI
apyI[dcm+]
mspI
nlaIII
nlaIV
hpaII
pciI
mboII
ddel sau96I|dcm-|
dsaV
nspHI
mnlI
bpuAI
eco81I avaII|dcm-)
nlaIV banII
nspI
mwoI
bsII
nlaIV
bbsI
bsmI bsu36I/mstII/sauI
cac8I
bssKI
aflIII
1701 GCTTGTCCTG CTGACCCCAG CAGACCCTCC TGTTGGTTCC AGCGAAGACC ACAGGCATTC CTTAGGGTCC AGGGTCAGCA GGCTCCGGGC TCACATGTGT
CGAACAGGAC GACTGGGGTC GTCTGGGAGG ACAACCAAGG TCGCTTCTGG TGTCCGTAAG GAATCCCAGG TCCCAGTCGT CCGAGGCCCG AGTGTACACA

```
                                                       mspI|M.haeIII-]
                                                       hpaII
                                                      naeI/ngoMI|M.haeIII-]
                                                  bpmI/gsuI|dcm-]
              tspRI                              scrFI|dcm-]
              alwNI|dcm-]                        pspGI   cfr10I/bsrFI                 aluI      dsaI
              alw26I/bsmAI                       mvaI    cac8I                       pflMI      btgI/
             mwoI                                ecoRII|dcm-]                        bslI       bsaJI
             btsI                                dsaV|dcm-]                         cac8I      haeIII/
             bstAPI                              bstNI  haeIII/palI               tseI pvuII   mscI/bal
            hpyCH4V                              bssKI|dcm-]                      fnu4HI/bsoFI eaeI
           sfcI                                  apyI|dcm+]                       bbvI mspA1I/nspBII
           pstI    cac8I        aluI  nlaIV    mnlI    ddeI   mboII     hphI     hpyCH4V      cfrI bce
1801 AAGGACAGGA CATTTTCTGC AGTGCCTGCC AATAGTGAGC TTGGAGCCTG GAGGCCGGCT TAGTTCTTCC ATTTCACCCT TGCAGCCAGC TGTTGGCCAC
     TTCCTGTCCT GTAAAAGACG TCACGGACGG TTATCACTCG AACCTCGGAC CTCCGGCCGA ATCAAGAAGG TAAAGTGGGA ACGTCGGTCG ACAACCGGTG

                                                  cac8I
     sau96I|M.haeIII-]                  mnlI   tseI             mnlI          mnlI
     nlaIV    mnlI          hpyCH4V    bsp1286   fnu4HI/bsoFI   ddeI          mnlI                 fokI
     haeIII/palI        bceAI sfaNI    bmyI      bbvI           bspCNI        bseRI    sfcI        bstF5I
1901 GGCCCCTGCC CTCTGGTCTG CCGTGCATCT CCCTGTGCCC TCTTGCTGCC TGCCTGTCTG CTGAGGTAAG GTGGGAGGAG GGCTACAGCC CACATCCCAC
     CCGGGGACGG GAGACCAGAC GGCACGTAGA GGGACACGGG AGAACGACGG ACGGACAGAC GACTCCATTC CACCCTCCTC CCGATGTCGG GTGTAGGGTG
```

```
                                                                  haeIII/palI
                                                                 stuI|dcm-]
                                                                scrFI|dcm-]
                                                                pspGI
                                                                mvaI
                                                                ecoRII|dcm-]
                                                    taiI        dsaV[dcm-]
                                                    maeII/hpyCH4IV
                                                    bsaXI       bstNI
                                                    pmlI        bssKI|dcm-)
                                                    eco72I mnlI apyI|dcm+]                    mwoI
                                                    bsaAI mwoI bsaJI                          bglI
                                                 tsp45I  sfiI|dcm-]                           hinlI/acyI
                                               mnlI  bbrPI bglI|M.haeIII-]                    hgaI
      mnlI                  nlaIII     hphI  maeIII  haeIII/palI                sfcI        sbaII/bsaHI
2001 CCCCTCGTCC AATCCCATAA TCCATGATGA AAGGTGAGGT CACGTGGCCT CCCAGGCCTG ACTTCCCAAC CTACAGCATT GACGCCAACT TGGCTGTGAA
     GGGGAGCAGG TTAGGGTATT AGGTACTACT TTCCACTCCA GTGCACCGGA GGGTCCGGAC TGAAGGGTTG GATGTCGTAA CTGCGGTTGA ACCGACACTT
```

```
                                                                                                      scr
                                                                                                      psp
                                                                                                      mva
                                                                                                      eco
            sau3AI                                                                                    dsa
            mboI/ndeII|dam-]             bsp1286                                                      bst
            dpnII|dam-]                  bmyI                          nlaIII               avaI      bss
            dpnI|dam+]        bsrI       ddeI                          sphI              bsp1286      apy
      mnlI  bstYI/xhoII       tsp45I     bspCNI             bsp1286    nspHI          bcgI bmyI     tseI
    mboII   alwI|dam-]        maeIII     hpy188I bslI       bmyI       nspI         hpyCH4V mwoI    fnu4HI/
    earI/ksp632I   haeIII/palI tspRI sfaNI banII             banII avaI cac8I        bsgI   banII    bbvI
2101 GGAAGAGGAA AGGATCTGGC CTTGTGGTCA CTGGCATCTG AGCCCTGCTG ATGGCTGGGG CTCTCGGGCA TGCTTGGGAG TGCAGGGGGC TCGGGCTGCC
     CCTTCTCCTT TCCTAGACCG GAACACCAGT GACCGTAGAC TCGGGACGAC TACCGACCCC GAGAGCCCGT ACGAACCCTC ACGTCCCCCG AGCCCGACGG

                                                                tth111I/aspI
                                                               sau96I|dcm-]   sau96I|M.haeIII-)
        tseI                                                   nlaIV          nlaIV
        fnu4HI/bsoFI                                           avaII|dcm-]    haeIII/palI
        bbvI                                                 scrFI|dcm-]     sau96I|M.haeIII-]
     scrFI|dcm-]                                             pspGI           pspOMI/bsp120I
     pspGI   mwoI                                            mvaI            nlaIV
     mvaI bsgI                                               ecoRII|dcm-]    eco0109I/draII
     ecoRII|dcm-]                                     ddeI|M.aluI-] pflFI    bsp1286|M.haeIII-)
     dsaV|dcm-]                                  hgiAI/aspHI  dsaV|dcm-)     bmyI        hpy188I
     bstNI hpyCH4V                        mwoI   bsp1286   cac8I bstNI       banII|M.haeIII-]
     bssKI|dcm-]                      bsp1286    bsiHKAI  aluI   bssKI|dcm-]   fokI apaI     ddeI             bsmF
     apyI|dcm+)                       bmyI       bmyI  bspCNI    apyI|dcm+)    bstF5I        bspCNI          bst4CI/
    haeIII/palI                       banII nlaIII   hpy188I  haeIII/palI   hpy188III       mnlI            tspRI bs
2201 TGGCCTGGCT GCACAGAAGG CAAGTGCTGG GGCTCATGGT GCTCTGAGCT GGCCTGGACC CTGTCAGGAT GGGCCCCACC TCAGAACCAA ACTCACTGTC
     ACCGGACCGA CGTGTCTTCC GTTCACGACC CCGAGTACCA CGAGACTCGA CCGGACCTGG GACAGTCCTA CCCGGGGTGG AGTCTTGGTT TGAGTGACAG
```

```
                              hgiAI/aspHI                                                                              tsp
                   mnlI tspRI bsp1286                                                                              haeIII/pa
       bst4CI/hpyCH4III       bsiHKAI                                                                              stuI tru9I
      tspRI     nlaIII btsI   bmyI  nlaIII  mnlI                                  mnlI                 mnlI  mseI
2301 CCCACTGTGG CATGAGGGCA GTGGAGCACC ATGTTTGAGG GCGAAGGGCA GAGCGTTTGT GTGTTCTGGG GAGGGAAGGA AAAGGTGTTG GAGGCCTTAA
     GGGTGACACC GTACTCCCGT CACCTCGTGG TACAAACTCC CGCTTCCCGT CTCGCAAACA CACAAGACCC CTCCCTTCCT TTTCCACAAC CTCCGGAATT

                                                                                     mboII
                                                                                     bpuAI
                                                                                     bbsI
                                                  mspI                              mnlI
        bst4CI/hpyCH4III         hpy188III        hpaII        bcgI         hpy188III      mnlI       aluI
2401 TTATGGACTG TTGGGAAAAG GGTTTTGTCC AGAAGGACAA GCCGGACAAA TGAGCGACTT CTGTGCTTCC AGAGGAAGAC GAGGGAGCAG GAGCTTGGCT
     AATACCTGAC AACCCTTTTC CCAAAACAGG TCTTCCTGTT CGGCCTGTTT ACTCGCTGAA GACACGAAGG TCTCCTTCTG CTCCCTCGTC CTCGAACCGA

                              scrFI|dcm-]
                              pspGI
                              mvaI
                              ecoRII|dcm-]                             sau96I
         pleI                 dsaV|dcm-]                               avaII
         mlyI                 bstNI                                  mwoI
         hinfI                bssKI|dcm-]                          mspAII/nspBII
       alwNI|dcm-]            bsaJI                               tseI  nlaIV
       alw26I/bsmAI        hinlI/acyI                             fnu4HI/bsoFI                      tspRI
        hpy188I            hgaI apyI|dcm+]                        bbvI  bsmFI                       btsI
       ddeI                ahaII/bsaHI               bslI        hinPI aciI                        tsp45I
       bspCNI        hpy188I bsaJI   nlaIV           bslI        hhaI/cfoI                         maeIII     nlaI
2501 GACTGCTCAG AGTCTGTTCT GACGCCCTGG GGGTTCCTGT CCAACCCCAG CAGGGGCGCA GCGGGACCAG CCCCACATTC CACTTGTGTC ACTGCTTGGA
     CTGACGAGTC TCAGACAAGA CTGCGGGACC CCCAAGGACA GGTTGGGGTC GTCCCCGCGT CGCCCTGGTC GGGGTGTAAG GTGAACACAG TGACGAACCT
```

```
                                                                    tru9I
                                                                    mseI
                                                                    aseI/asnI/vspI
                                                                   tsp509I
                                                                tru9I
                                                                pacI
                                                                mseI     aluI                              ms
2601 ACCTATTTAT TTTGTATTTA TTTGAACAGA GTTATGTCCT AACTATTTTT ATAGATTTGT TTAATTAATA GCTTGTCATT TTCAAGTTCA TTTTTTATTC
     TGGATAAATA AAACATAAAT AAACTTGTCT CAATACAGGA TTGATAAAAA TATCTAAACA AATTAATTAT CGAACAGTAA AAGTTCAAGT AAAAAATAAG

                                                                                     sfcI
                                                                                     pstI[M.H1-]
                                                                                    tseI
                                                                                 mspA1I/nspBII
                                                                                 aciI hpyCH4V
                                                                          styI haeIII/palI
                                                                          bsaJI fnu4HI/bsoFI
                                                        mnlI             haeIII/palI fnu4HI/bsoFI
                                           bslI  tspRI          bseRI           sau96I[M.haeIII-]  tsp45I
                         rsaI              aciI bstXI fokI   mnlI  mnlI        nlaIV   sau96I[M.haeIII-]
               nlaIII    csp6I           cac8I  bsrI  bstF5I bseRI bseRI       eco0109I/draII bbvI  maeIII     hpyI
2701 ATATTTATGT TCATGGTTGA TTGTACCTTC CCAAGCCCGC CCAGTGGGAT GGGAGGAGGA GGAGAAGGGG GGCCTTGGGC CGCTGCAGTC ACATCTGTCC
     TATAAATACA AGTACCAACT AACATGGAAG GGTTCGGGCG GGTCACCCTA CCCTCCTCCT CCTCTTCCCC CCGGAACCCG GCGACGTCAG TGTAGACAGG
```

nlaIV
scrFI|dcm-|
pspGI
mvaI
ecoRII|dcm-|
tseI dsaV|dcm-|
mwoI bstNI
haeIII/palI fnu4HI/bsoFI
eaeI tseI bssKI|dcm-|
bpmI/gsuI|dcm-| cfrI fnu4HI/bsoFI
eco8
tsp509I bsrI fnu4HI/bsoFI bbvI apyI|dcm+| ecoNI bsu3
apoI bsmFI mnlI aciI bbvI bsaJI bslI bspMI cac8I bslI dde
2801 AGAGAAATTC CTTTTGGGAC TGGAGGCAGA AAAGCGGCCA GAAGGCAGCA GCCCTGGCTC CTTTCCTTTG GCAGGTTGGG GAAGGGCTTG CCCCCAGCCT
TCTCTTTAAG GAAAACCCTG ACCTCCGTCT TTTCGCCGGT CTTCCGTCGT CGGGACCGAG GAAAGGAAAC CGTCCAACCC CTTCCCGAAC GGGGGTCGGA
tseI
fnu4HI/bsoFI
bsrI bbvI
mnlI mnlI tfiI hinPI mnlI
bsrI mnlI nlaIV hinfI hhaI/cfoI
2901 TAGGATTTCA GGGTTTGACT GGGGGCGTGG AGAGAGAGGG AGGAACCTCA ATAACCTTGA AGGTGGAATC CAGTTATTTC CTGCGCTGCG AGGGTTTCTT
ATCCTAAAGT CCCAAACTGA CCCCCGCACC TCTCTCTCCC TCCTTGGAGT TATTGGAACT TCCACCTTAG GTCAATAAAG GACGCGACGC TCCCAAAGAA
nlaIV tsp509I mnlI mnlI
tspRI banI bst4CI/hpyCH4IV bseRI nlaIV mn
hpy188I btsI mnlI mnlI tspRI apoI aciI bpmI/gsuI|dcm-| cac8I bsmFI
3001 TATTTCACTC TTTTCTGAAT GTCAAGGCAG TGAGGTGCCT CTCACTGTGA ATTTGTGGTG GGCGGGGGCT GGAGGAGAGG GTGGGGGGCT GGCTCCGTCC
ATAAAGTGAG AAAAGACTTA CAGTTCCGTC ACTCCACGGA GAGTGACACT TAAACACCAC CCGCCCCCGA CCTCCTCTCC CACCCCCCGA CCGAGGCAGG

```
                                            haeIII/palI
                                            eaeI
                                            cfrI
                                            mspI|M.haeIII-]
              tseI                          hpaII
              mkoI         tru9I            naeI/ngoMI|M.haeIII-]
              fnu4HI/bsoFI                  cfr10I/bsrFI               hpyCH4V
              bbvI         mseI             cac8I     bsrI      mnlI bst4CI/hpyCH4III   mslI              mnlI      mboII
3101 CTCCCAGCCT TCTGCTGCCC TTGCTTAACA ATGCCGGCCA ACTGGCGACC TCACGGTTGC ACTTCCATTC CACCAGAATG ACCTGATGAG GAAATCTTCA
     GAGGGTCGGA AGACGACGGG AACGAATTGT TACGGCCGGT TGACCGCTGG AGTGCCAACG TGAAGGTAAG GTGGTCTTAC TGGACTACTC CTTTAGAAGT

              sau3AI
        hpyCH4V        hpyCH4V                               pleI
       sfaNI    mboI/ndeII(dam-)                             mlyI               tru9I            taiI
      fokI      dpnII(dam-)                                hpmI/gsuI(dcm-]      mseI             maeII/hpyCH4IV
      bstF5I    dpnI|dam+]   tsp509I                     bsrI hinfI           tsp509I    tsp509I                   ah
3201 ATAGGATGCA AAGATCAATG CAAAAATTGT TATATATGAA CATATAACTG GAGTCGTCAA AAAGCAAATT AAGAAAGAAT TGGACGTTAG AAGTTGTCAT
     TATCCTACGT TTCTAGTTAC GTTTTTAACA ATATATACTT GTATATTGAC CTCAGCAGTT TTTCGTTTAA TTCTTTCTTA ACCTGCAATC TTCAACAGTA

        tseI
        fnu4HI/bsoFI
     tru9I
     mseI bbvI                          aluI
3301 TTAAAGCAGC CTTCTAATAA AGTTGTTTCA AAGCTGAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
     AATTTCGTCG GAAGATTATT TCAACAAAGT TTCGACTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT
                                                                                                     end insert^
```

```
                                    scrFI|M.hpaII-]
                                    nciI
                                    mspI
                                    hpaII
                                   xmaI/pspAI          cfr10I/bsrFI
                                   smaI                sau96I
                                   scrFI|M.hpaII-]     rsrII/cspI
                                   nciI                cpoI
          aciI                     dsaV               mspI    nlaIV
         fnu4HI/bsoFI   alwI       bssKI              hpaII   kpnI sfcI
        haeIII/paeI     sstI   salI bsaJI             mroI bsaWI   sse8387I
       mcrI             sacI   bincII/hindII|M.taqI-| avaII|M.hpaII-]
       eagI/xmaIII/eclXI       accI|M.taqI-] tru9I hpy188III rsaI hpyCH4V
       eaeI             hgiAI/aspHI|M.aluI-] mseI  bspMII   banI pstI
       cfrI       rmaI ecl136II    avaI|M.hpaII-| bspEI mspI   sbfI
       bsiEI      maeI bsp1286|M.aluI-)  xmnI tsp509I  ageI csp6I
      notI        bfaI bsiHKAI taqI  dsaV tsp509I bsaWI hpaII bspMI
      fnu4HI/bsoFI    bmyI  hpy99I  bssKI aseI/asnI/vspI  asp718        rsaI aluI
      aciI       speI  banII|M.aluI-]    asp700   accIII   acc65I cac8I  csp6I
3401 GGGCGGCCGC CGACTAGTGA GCTCGTCGAC CCGGGAATTA ATTCCGGACC GGTACCTGCA GGCGTACCAG CTTTCCC
     CCCGCCGGCG GCTGATCACT CGAGCAGCTG GGCCCTTAAT TAAGGCCTGG CCATGGACGT CCGCATGGTC GAAAGGG
```

> length: 3477

aatII(GACGTC):	24
acc65I(GGTACC):	3451
accI(GTMKAC):	1192 1492 3425
accIII(TCCGGA):	1368 3443

EXHIBIT B

[redacted]

User:grim

oli.out3_6_2000

Tue Feb 1 11:45:17 PST 2005

xe12380 / XE12380

[redacted]

```
xe12380  User:grim Job: oli.out3_6_2000 Date: Tue Feb  1 11:45:17 PST 2005

xe12380  User:grim Job: oli.out3_6_2000 Date: Tue Feb  1 11:45:17 PST 2005

xe12380  User:grim Job: oli.out3_6_2000 Date: Tue Feb  1 11:45:17 PST 2005

xe12380  User:grim Job: oli.out3_6_2000 Date: Tue Feb  1 11:45:17 PST 2005
```

<16422.f10 OLI28176>AGTTGCAACTATGCTGGCAACTGGAGG
<16422.r10 OLI28177>CCCAGATGTTTCTTACTGGCTCCTCACTAATC
<16435.f10 OLI28178>GAGGCAACTAAAAAGGCTTCAAACGTTTTG
<16435.r10 OLI28179>CAAAATCAATAACGTCATCAGCTTCCTAACCATG
<23334.f10 OLI28180>CGTTGGTTGAAGGACCTAAATACCTGGC
<23334.r10 OLI28181>CTTCTATCATCACCCAGCTGCATGACC
<26843.f10 OLI28182>GCAAGTTCATGCTCTGAGTCCTGAAGAG
<26843.r10 OLI28183>GGTGCAGTGTCATAGAGGGTTAATCCATG
<26844.f10 OLI28184>CCAATTCTGAACATTCCCATCGTGC
<26844.r10 OLI28185>GTGCTGGGATTACAGGCGTGAGC
<26844.f10>CCAATTCTGAACATTCCCATCGTGC
<26844.r10>GTGCTGGGATTACAGGCGTGAGC
<30862.f10 OLI28186>CACCTGAAGGTGATGCTCCTGGAAG
<30862.r10 OLI28187>GCGGTTCCATCGCAGATGACC
<40621.f10 OLI28188>CTGGCTGCCCATCATGACCTCC
<40621.r10 OLI28189>CTCCTGCCTCAGCCTCCCGAGTAG
<44161.f10 OLI28190>GAAGGACATGCGCGTGCAGAC
<44161.r10 OLI28191>GAGCTGGTGCTTTCCGCTGCC
<44694.f10 OLI28192>CCACAGCTGATACGGCATCCTGC
<44694.r10 OLI28193>CCAGAGTCTGCATGAGCACCAATG
<48320.f10 OLI28194>CCTCACTAGCACCTGGAATGATGCTTTG
<48320.r10 OLI28195>CACTCATAAGTTGCACATATGCTCCAAGGTC
<48334.f10 OLI28196>CAAGGAGGCAAAATTTTGACAGGGAAGG
<48334.r10 OLI28197>CTACACACAGAATAAGGTTGGGGAATTAAGCTG
<48606.f10 OLI28198>CAGCATGCCAGGCCTCACG
<48606.r10 OLI28199>CTTTCTCAGGCCTCCTGGCCAATAG
<49141.f10 OLI28200>CCACGTATCTCATGTGCCGAATGTG
<49141.r10 OLI28201>GTGCAGCCTCACACTGCCTTCTCC
<49647.f10 OLI28202>GGCAGTGAACACATCTGATTTCCCACAG
<49647.r10 OLI28203>CCCTCAAAAGCCAGTTACATTACATGTTCACAG
<49819.f10 OLI28204>GGTTGCCATTCCATAGGTTTGGAGAGC
<49819.r10 OLI28205>GCATTCACCTACCTGTCTAACCCTCACATG
<49820.f10 OLI28206>GAGAAGCTGACTGAGGAAGGCTCTCCC
<49820.r10 OLI28207>CTACTGGAAGCAGGCACAGTGTCACTAGC
<53913.f10 OLI28208>GCTTCCGTCCCTGAATCCCTTCC
<53913.r10 OLI28209>GTGTGAAACTGCTTGGTGGCTGTTCC
<53978.f10 OLI28210>GTGATGAATGGTCACACACCGATGCAC
<53978.r10 OLI28211>GAATCCTGGGGGATCAAATCATATGAATG
<53996.f10 OLI28212>CCAGATGGCACGACACTGCCATG
<53996.r10 OLI28213>CCAGAACAGCAGTGAGGCAGATTGATG
<56050.f10 OLI28214>GAAGGCTTCCTGCAGGAGCAGTCTG
<56050.r10 OLI28215>CAGGCCAGGTGAGAGAGTACAAGTCTTGC
<56110.f10 OLI28216>GTAGCCGGCTTGGCTTTTGTAATTGG
<56110.r10 OLI28217>GGATGGCTGAGTTTCGCACATTGTG
<56410.f10 OLI28218>CCTGAAGCTAGTCCAGCTAGTACACCACAAATC
<56410.r10 OLI28219>CAATCCTCCTGACTTGTCTTCCCTAAGTGC
<56436.f10 OLI28220>GAAATGGATTAGCCAACCAGGGCAAC
<56436.r10 OLI28221>GGAGCAGGAACCAACTCAATGCACAG
<56855.f10 OLI28222>GCAACTTACAGCTGCACCGACAGTTG
<56855.r10 OLI28223>GAGAGCACTGGAATGATTTAGGGGTGG
<56859.f10 OLI28224>GGCTGTCTTAGTACTTCGCCTGACAGTTGTC
<56859.r10 OLI28225>CCATCAACTATGTGATCCCAAAGCGC
<56860.f10 OLI28226>CTTCATGTGGCAGCAAGTTTTCGAGC
<56860.r10 OLI28227>GAGACGCATGGGTGTAGCCTCAGG
<56865.f10 OLI28228>CTAACCAACTGGAAAAATGGATGAAACTCAATG
<56865.r10 OLI28229>GTCTGTCTGGGAAGCTGACTGCCAGAC
<56868.f10 OLI28230>GGATTCCGAGAATCATTGTCATGTACATGC
<56868.r10 OLI28231>GTTCCCTCCTCATTCCTTAATGAGTGCTTG
<56869.f10 OLI28232>CAACTTCGCCCTGGAGCAGCTCTTC
<56869.r10 OLI28233>CTCCAGCACCCTCAACATGATGTAGCC

<56870.f10 OLI28234>GAACTCGGAAGAGATGCTGTGGAATCC
<56870.r10 OLI28235>CCTGTCCCAGTAATGCCAACTTGGAG
<57699.f10 OLI28236>CCTACGGTGAGAAGCTTACCATAAGCTTGG
<57699.r10 OLI28237>CATGGCATGGCTACCGCTTCCTTAG
<57704.f10 OLI28238>GCCCCAATGCCAAGTACGTAATGAAG
<57704.r10 OLI28239>CAGTGTAAGCCAGCACACTGACCTCC
<57710.f10 OLI28240>CAGCCTCCAAGCCATCATCACCAAG
<57710.r10 OLI28241>GTCCTGAGCGACTGCCACCATG
<57711.f10 OLI28242>GCATCGAAGCCCGTGAAATCCAG
<57711.r10 OLI28243>GTCAGGCACTGTGGTGAAGGGAACG
<57827.f10 OLI28244>GCCATCAATGGACATGATCTTCGATATGG
<57827.r10 OLI28245>CCATGTTGTGGTTGGAGTCCAGGG
<57844.f10 OLI28246>CCGTGGAATGGAGTTGATCCCAACC
<57844.r10 OLI28247>GCCTATCTAAGGTCCCTTGCTGCAAGG
<58723.f10 OLI28248>GATAAAACATGCAGCCACAGGCTCTCC
<58723.r10 OLI28249>GGACCAGACTGTACTGTGGCCATGTACAC
<58737.f10 OLI28250>GCACACATTAAGAACCTGTTACAGCTCATTGTTG
<58737.r10 OLI28251>GAAGGCTATCAGCACCTGCAACCAGC
<58743.f10 OLI28252>CACATGCGTCACATGGGCATTTC
<58743.r10 OLI28253>GTTCGCACCATTCTCCTGCCTCAG
<58846.f10 OLI28254>CCCAGCAGTGGGACAGCCAGAC
<58846.r10 OLI28255>CAAGACCTATGTTCTGGGGCAGCAGG
<58848.f10 OLI28256>GAACTCCCCACCTTTGCACGCTG
<58848.r10 OLI28257>GATGTGTCCTTCACTCACCCGCAGC
<58849.f10 OLI28258>CGAACACCTTGTACTGGGAGTTGAATCAG
<58849.r10 OLI28259>CATCACATAGCAGAGTTCCCTCAGCCCTG
<58850.f10 OLI28260>GCACCTGTCAAAGCCTAAAGTCACCATG
<58850.r10 OLI28261>CCACAGTGGAGTAAACCGTATTTGCTGG
<58853.f10 OLI28262>GGGACTGACCCTAGTCTGTGTCCATGC
<58853.r10 OLI28263>GGGATGCTGTATGGATAGGAAGGGATG
<58855.f10 OLI28264>CGAAGGATGGACATCCTGCAATGG
<58855.r10 OLI28265>GCTTAGAGACTCCACACAGACAGCCAAGG
<59211.f10 OLI28266>CAAGGTGACCTCGCAGGACACTGG
<59211.r10 OLI28267>CCTCCTGATCAGAAAGGGGCCTAACAG
<59213.f10 OLI28268>GAGGCTGACACCTTCATGTTTGGAGG
<59213.r10 OLI28269>CTCAGTGTGGTCTGGCAGGAACCG
<59497.f10 OLI28270>CCTGACCAAAAAATTCCCAGTAACCAGGC
<59497.r10 OLI28271>GTGGCCAAGTGGATAAAACAGTAGCAGTG
<59603.f10 OLI28272>CTGCTGTGGCTGCAGCTCTGC
<59603.r10 OLI28273>GACGCCAGGAAAACAGCCAGGTC
<59605.f10 OLI28274>GGAGTACCATCTTCCTCATGGGACCAG
<59605.r10 OLI28275>CATGGGAATTCAGTGGGACAACATCC
<59607.f10 OLI28276>CCACTACATGAGCATCACCATCTTGGTC
<59607.r10 OLI28277>CTGGCCAGTGACTGTCAGGTTCTCTG
<59609.f10 OLI28278>CATTCCACTTGTGTCACTGCTTGGAACC
<59609.r10 OLI28279>GGCATTGTTAAGCAAGGGCAGCAGAAG
<59610.f10 OLI28280>CTGCCTCTCAGCCCTTACCTGATGC
<59610.r10 OLI28281>CCCTGAGTTGCGAAGTGGCAGTC
<59612.f10 OLI28282>GTGAGCGGATGGACCTAGCACTTCC
<59612.r10 OLI28283>GACAGCTACTACTCGCCAGGTGTGCTG
<59613.f10 OLI28284>GAATGCAGCAACACCAGCGATGC
<59613.r10 OLI28285>GACCTGTGACTCCCACTTACTGGGCAG
<59616.f10 OLI28286>GTCAGGAGCCTGAAATCAGGCTGACTTC
<59616.r10 OLI28287>TCCTCCAGGCCAATCCTGACCC
<59619.f10 OLI28288>GTGCATGTTCCCTGTCTGGTGCC
<59619.r10 OLI28289>GAGTCCCTGAGGTTGGAGTCCTAGCATAGC
<59625.f10 OLI28290>GTATTCTGCCCTGCGCATCCCAC
<59625.r10 OLI28291>GACTCCTCCAGTGGCACTGCAGCTC
<59827.f10 OLI28292>CACTGCAGCTGCGCACAGTCG
<59827.r10 OLI28293>GGTGCTCTGAGAAGAGGTCAGAATGGC

<59828.f10 OLI28294>CTGCCGACACCGGACTTCAGC
<59828.r10 OLI28295>GCCAGCCGACCTTTCTGTGGTG
<59853.f10 OLI28296>CTCCTTAGCCCGTGTGAGCCTCAC
<59853.r10 OLI28297>GTGCAATCTCAGCTCACTGCAGCCTC
<59854.f10 OLI28298>GGAGAGGCCACCGGGACTTCAG
<59854.r10 OLI28299>GGGGCCTGCTGAAAGACAGGGTC
<60283.f10 OLI28300>CTGTTTCTGATGTGGGGTTCCTCCAC
<60283.r10 OLI28301>CCATTGGCTTTCAAGGACCTTTCCCAAC
<60619.f10 OLI28302>CAGGCCCAGTGGGTGGAATCTGTC
<60619.r10 OLI28303>CATCTCCCGAAGAGAGAAGTCCCAACC
<60625.f10 OLI28304>CCAGGAACCTGAGCTAGGTCAAAGACG
<60625.r10 OLI28305>GTGAATAGCCCCATCCTCTCTTGACCTC
<60629.f10 OLI28306>TGATCTGTCACCAAAGGAGGCCAACTC
<60629.r10 OLI28307>CAGGTATCAGTATGAAAAAGGATCTTTGTTCATCAC
<61755.f10 OLI28308>GAATGAGCCTATCAGCAGGGCTCTAGTTTC
<61755.r10 OLI28309>GGAAGCCAGAAGCCAGGAGGAGC
<64852.f10 OLI28310>CAAGCCCTGCAGACAGTCTGTCTCC
<64852.r10 OLI28311>GGCTGATGCTATCATATTGCTATGGACATTG
<66308.f10 OLI28312>CGACATCCAGGCCAAGGACCTG
<66308.r10 OLI28313>CATCCTTGGGATCTCATGGTTGGG
<68869.f10 OLI28314>GCGTACTCCACAGTGCAGAGAGTCGC
<68869.r10 OLI28315>GGCTATCCTAAGTACCTCATGAAGAGGAGGC

EXHIBIT C

Normal vs. Tumor Library Set #1

6/13/00

	DNA #	821-823 H. Esophagus	762-763 H. Esophagus Tum.	798-799 H. Stomach	786-787 H. Stomach Tumor	792-793 H. Kidney	824-825 H. Kidney Tumor	794-795 H. Lung	776-778 H. Lung Tumor	796-797 H. Rectum	784-785 H. Rectum Tumor	833-834 H. Skin	764-766 H. Melanoma	expected size bp	source for cloning
1	58850	+−	++	++	+	+−	++	++	+	++	++	+−	−		
2	58853	−	−	−	−	−	−	−	−	−	−	+−	−		
3	58855	−	−	++	+−	−	−	−	−	+−	++	−	−		
4	59211	−	−	−	−	++	−	−	−	−	−	−	−		
5	59213	+−	+−	+	+−	+−	+−	+−	+−	++	+	+	−	531	
6	59497	+	++	++	+	+	++	+	+	+	+	+	++		
7	59603	−	−	−	−	−	−	−	−	−	−	−	−		
8	59605	−	−	−	−	−	−	−	−	−	−	−	++		
9	59607	−	−	−	−	−	−	−	−	−	−	−	−		
10	59609	−	+−	−	−	−	−	−	−	+−	+−	+−	+−		
11	59610	−	+	+−	−	++	+	−	++	−	+−	+	−		
12	59612	+−	+−	+−	+−	+−	+−	+−	+−	+−	+−	+−	+−		
13	59613	−	−	−	−	−	−	−	−	++	++	−	−		
14	59616	+−	+−	+−	+−	+−	+−	+−	+−	+−	+−	+−	+−		
15	59619	−	−	−	−	−	−	−	−	−	+−	−	−		
16	59625	+−	+−	+−	+−	+−	+−	+−	+−	+−	+−	+−	+−		
17															

59609

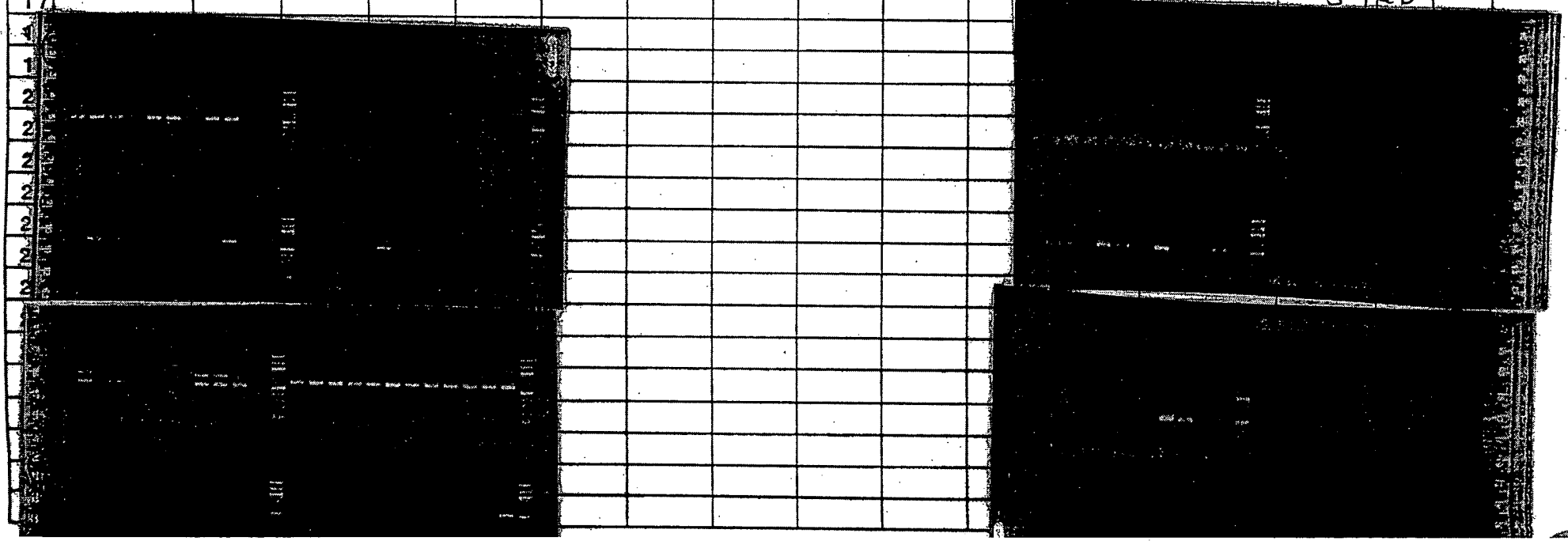


EXHIBIT 1

MOLECULAR BIOLOGY OF THE CELL

THIRD EDITION

Text Editor: Miranda Robertson
Managing Editor: Ruth Adams
Illustrator: Nigel Orme
Molecular Model Drawings: Kate Hesketh-Moore
Director of Electronic Publishing: John M-Roblin
Computer Specialist: Chuck Bartelt
Disk Preparation: Carol Winter
Copy Editor: Shirley M. Cobert
Production Editor: Douglas Goertzen
Production Coordinator: Perry Bessas
Indexer: Maija Hinkle

Bruce Alberts received his Ph.D. from Harvard University and is currently President of the National Academy of Sciences and Professor of Biochemistry and Biophysics at the University of California, San Francisco. *Dennis Bray* received his Ph.D. from the Massachusetts Institute of Technology and is currently a Medical Research Council Fellow in the Department of Zoology, University of Cambridge. *Julian Lewis* received his D.Phil. from the University of Oxford and is currently a Senior Scientist in the Imperial Cancer Research Fund Developmental Biology Unit, University of Oxford. *Martin Raff* received his M.D. from McGill University and is currently a Professor in the MRC Laboratory for Molecular Cell Biology and the Biology Department, University College London. *Keith Roberts* received his Ph.D. from the University of Cambridge and is currently Head of the Department of Cell Biology, the John Innes Institute, Norwich. *James D. Watson* received his Ph.D. from Indiana University and is currently Director of the Cold Spring Harbor Laboratory. He is the author of *Molecular Biology of the Gene* and, with Francis Crick and Maurice Wilkins, won the Nobel Prize in Medicine and Physiology in 1962.

Library of Congress Cataloging-in-Publication Data
Molecular biology of the cell / Bruce Alberts . . . [et al.].—3rd ed.
p. cm.
Includes bibliographical references and index.
ISBN 0-8153-1619-4 (hard cover).—ISBN 0-8153-1620-8 (pbk.)
1. Cytology. 2. Molecular biology. I. Alberts, Bruce.
[DNLM: 1. Cells. 2. Molecular Biology. QH 581.2 M718 1994]
QH581.2.M64 1994
574.87—dc20
DNLM/DLC
for Library of Congress 93-45907
CIP

Published by Garland Publishing, Inc.
717 Fifth Avenue, New York, NY 10022

Printed in the United States of America

15 14 13 12 10 9 8 7

Front cover: The photograph shows a rat nerve cell in culture. It is labeled (*yellow*) with a fluorescent antibody that stains its cell body and dendritic processes. Nerve terminals (*green*) from other neurons (not visible), which have made synapses on the cell, are labeled with a different antibody. (Courtesy of Olaf Mundigl and Pietro de Camilli.)

Dedication page: Gavin Borden, late president of Garland Publishing, weathered in during his mid-1980s climb near Mount McKinley with MBoC author Bruce Alberts and famous mountaineer guide Mugs Stump (1940–1992).

Back cover: The authors, in alphabetical order, crossing Abbey Road in London on their way to lunch. Much of this third edition was written in a house just around the corner. (Photograph by Richard Olivier.)

extracts. If these minor cell proteins differ among cells to the same extent as the more abundant proteins, as is commonly assumed, only a small number of protein differences (perhaps several hundred) suffice to create very large differences in cell morphology and behavior.

A Cell Can Change the Expression of Its Genes in Response to External Signals [3]

Most of the specialized cells in a multicellular organism are capable of altering their patterns of gene expression in response to extracellular cues. If a liver cell is exposed to a glucocorticoid hormone, for example, the production of several specific proteins is dramatically increased. Glucocorticoids are released during periods of starvation or intense exercise and signal the liver to increase the production of glucose from amino acids and other small molecules; the set of proteins whose production is induced includes enzymes such as tyrosine aminotransferase, which helps to convert tyrosine to glucose. When the hormone is no longer present, the production of these proteins drops to its normal level.

Other cell types respond to glucocorticoids in different ways. In fat cells, for example, the production of tyrosine aminotransferase is reduced, while some other cell types do not respond to glucocorticoids at all. These examples illustrate a general feature of cell specialization—different cell types often respond in different ways to the same extracellular signal. Underlying this specialization are features that do not change, which give each cell type its permanently distinctive character. These features reflect the persistent expression of different sets of genes.

Gene Expression Can Be Regulated at Many of the Steps in the Pathway from DNA to RNA to Protein [4]

If differences between the various cell types of an organism depend on the particular genes that the cells express, at what level is the control of gene expression exercised? There are many steps in the pathway leading from DNA to protein, and all of them can in principle be regulated. Thus a cell can control the proteins it makes by (1) controlling when and how often a given gene is transcribed (**transcriptional control**), (2) controlling how the primary RNA transcript is spliced or otherwise processed (**RNA processing control**), (3) selecting which completed mRNAs in the cell nucleus are exported to the cytoplasm (**RNA transport control**), (4) selecting which mRNAs in the cytoplasm are translated by ribosomes (**translational control**), (5) selectively destabilizing certain mRNA molecules in the cytoplasm (**mRNA degradation control**), or (6) selectively activating, inactivating, or compartmentalizing specific protein molecules after they have been made (**protein activity control**) (Figure 9–2).

For most genes transcriptional controls are paramount. This makes sense because, of all the possible control points illustrated in Figure 9–2, only transcriptional control ensures that no superfluous intermediates are synthesized. In the

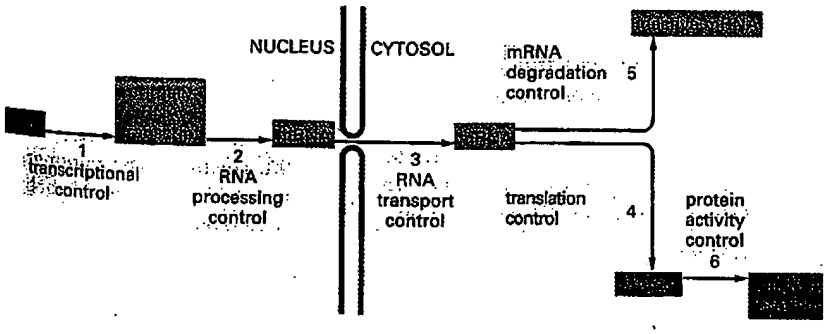


Figure 9–2 Six steps at which eucaryote gene expression can be controlled. Only controls that operate at steps 1 through 5 are discussed in this chapter. The regulation of protein activity (step 6) is discussed in Chapter 5; this includes reversible activation or inactivation by protein phosphorylation as well as irreversible inactivation by proteolytic degradation.

following sections we discuss the DNA and protein components that regulate the initiation of gene transcription. We return at the end of the chapter to the other ways of regulating gene expression.

Summary

The genome of a cell contains in its DNA sequence the information to make many thousands of different protein and RNA molecules. A cell typically expresses only a fraction of its genes, and the different types of cells in multicellular organisms arise because different sets of genes are expressed. Moreover, cells can change the pattern of genes they express in response to changes in their environment, such as signals from other cells. Although all of the steps involved in expressing a gene can in principle be regulated, for most genes the initiation of RNA transcription is the most important point of control.

DNA-binding Motifs in Gene Regulatory Proteins [5]

How does a cell determine which of its thousands of genes to transcribe? As discussed in Chapter 8, the transcription of each gene is controlled by a regulatory region of DNA near the site where transcription begins. Some regulatory regions are simple and act as switches that are thrown by a single signal. Other regulatory regions are complex and act as tiny microprocessors, responding to a variety of signals that they interpret and integrate to switch the neighboring gene on or off. Whether complex or simple, these switching devices consist of two fundamental types of components: (1) short stretches of DNA of defined sequence and (2) *gene regulatory proteins* that recognize and bind to them.

We begin our discussion of gene regulatory proteins by describing how these proteins were discovered.

Gene Regulatory Proteins Were Discovered Using Bacterial Genetics [6]

Genetic analyses in bacteria carried out in the 1950s provided the first evidence of the existence of **gene regulatory proteins** that turn specific sets of genes on or off. One of these regulators, the *lambda repressor,* is encoded by a bacterial virus, *bacteriophage lambda.* The repressor shuts off the viral genes that code for the protein components of new virus particles and thereby enables the viral genome to remain a silent passenger in the bacterial chromosome, multiplying with the bacterium when conditions are favorable for bacterial growth (see Figure 6–80). The lambda repressor was among the first gene regulatory proteins to be characterized, and it remains one of the best understood, as we discuss later. Other bacterial regulators respond to nutritional conditions by shutting off genes encoding specific sets of metabolic enzymes when they are not needed. The *lac repressor,* for example, the first of these bacterial proteins to be recognized, turns off the production of the proteins responsible for lactose metabolism when this sugar is absent from the medium.

The first step toward understanding gene regulation was the isolation of mutant strains of bacteria and bacteriophage lambda that were unable to shut off specific sets of genes. It was proposed at the time, and later proved, that most of these mutants were deficient in proteins acting as specific repressors for these sets of genes. Because these proteins, like most gene regulatory proteins, are present in small quantities, it was difficult and time-consuming to isolate them. They were eventually purified by fractionating cell extracts on a series of standard chromatography columns (see pp. 166–169). Once isolated, the proteins were shown to bind to specific DNA sequences close to the genes that they

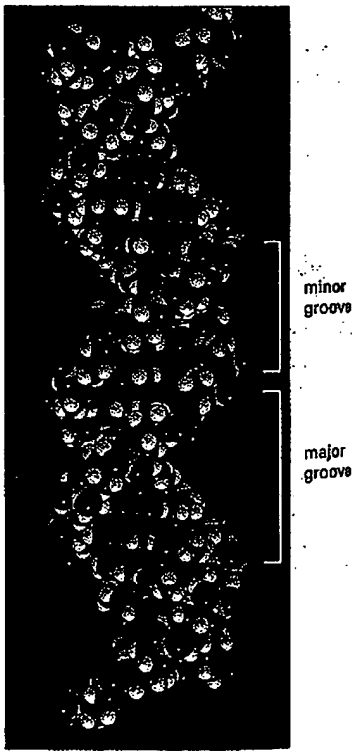


Figure 9–3 Double-helical structure of DNA. The major and minor grooves on the outside of the double helix are indicated. The atoms are colored as follows: carbon, *dark blue;* nitrogen, *light blue;* hydrogen, *white;* oxygen, *red;* phosphorus, *yellow.*

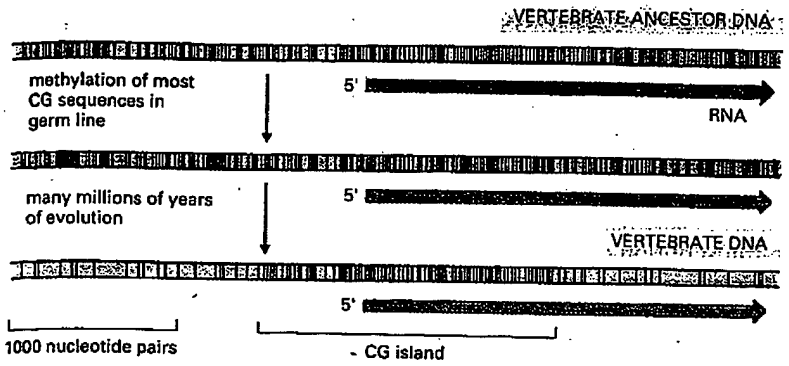


Figure 9–71 A mechanism to explain both the marked deficiency of CG sequences and the presence of CG islands in vertebrate genomes. A *black line* marks the location of an unmethylated CG dinucleotide in the DNA sequence, while a *red line* marks the location of a methylated CG dinucleotide.

Summary

The many types of cells in animals and plants are created largely through mechanisms that cause different genes to be transcribed in different cells. Since many specialized animal cells can maintain their unique character when grown in culture, the gene regulatory mechanisms involved in creating them must be stable once established and heritable when the cell divides, endowing the cell with a memory of its developmental history. Procaryotes and yeasts provide unusually accessible model systems in which to study gene regulatory mechanisms, some of which may be relevant to the creation of specialized cell types in higher eucaryotes. One such mechanism involves a competitive interaction between two (or more) gene regulatory proteins, each of which inhibits the synthesis of the other; this can create a flip-flop switch that switches a cell between two alternative patterns of gene expression. Direct or indirect positive feedback loops, which enable gene regulatory proteins to perpetuate their own synthesis, provide a general mechanism for cell memory.

In eucaryotes gene transcription is generally controlled by combinations of gene regulatory proteins. It is thought that each type of cell in a higher eucaryotic organism contains a specific combination of gene regulatory proteins that ensures the expression of only those genes appropriate to that type of cell. A given gene regulatory protein may be expressed in a variety of circumstances and typically is involved in the regulation of many genes.

In addition to diffusible gene regulatory proteins, inherited states of chromatin condensation are also utilized by eucaryotic cells to regulate gene expression. In vertebrates DNA methylation also plays a part, mainly as a device to reinforce decisions about gene expression that are made initially by other mechanisms.

Posttranscriptional Controls

Although controls on the initiation of gene transcription are the predominant form of regulation for most genes, other controls can act later in the pathway from RNA to protein to modulate the amount of gene product that is made. Although these **posttranscriptional controls,** which operate after RNA polymerase has bound to the gene's promoter and begun RNA synthesis, are less common than *transcriptional control,* for many genes they are crucial. It seems that every step in gene expression that could be controlled in principle is likely to be regulated under some circumstances for some genes.

We consider the varieties of posttranscriptional regulation in temporal order, according to the sequence of events that might be experienced by an RNA molecule after its transcription has begun (Figure 9–72).

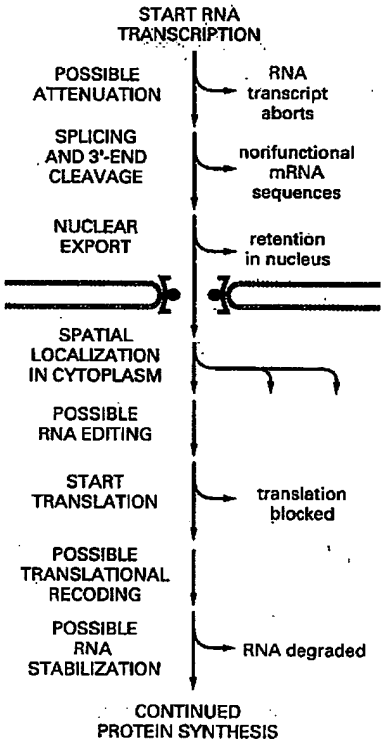


Figure 9–72 Possible posttranscriptional controls on gene expression. Only a few of these controls are likely to be used for any one gene.

EXHIBIT 2

CHAPTER 29

Regulation of transcription

Genes VII (1997) CH 29, pp. 847-848.
Benjamin Lewin

The phenotypic differences that distinguish the various kinds of cells in a higher eukaryote are largely due to differences in the expression of genes that code for proteins, that is, those transcribed by RNA polymerase II. In principle, the expression of these genes might be regulated at any one of several stages. The concept of the "level of control" implies that gene expression is not necessarily an automatic process once it has begun. It could be regulated in a gene-specific way at any one of several sequential steps. We can distinguish (at least) five potential control points, forming the series:

Activation of gene structure
↓
Initiation of transcription
↓
Processing the transcript
↓
Transport to cytoplasm
↓
Translation of mRNA

The existence of the first step is implied by the discovery that genes may exist in either of two structural conditions. Relative to the state of most of the genome, genes are found in an "active" state in the cells in which they are expressed (see Chapter 27). The change of structure is distinct from the act of transcription, and indicates that the gene is "transcribable." This suggests that acquisition of the "active" structure must be the first step in gene expression.

Transcription of a gene in the active state is controlled at the stage of initiation, that is, by the interaction of RNA polymerase with its promoter. This is now becoming susceptible to analysis in the *in vitro* systems (see Chapter 28). For most genes, this is a major control point; probably it is the most common level of regulation.

There is at present no evidence for control at subsequent stages of transcription in eukaryotic cells, for example, via antitermination mechanisms.

The primary transcript is modified by capping at the 5' end, and usually also by polyadenylation at the 3' end. Introns must be spliced out from the transcripts of interrupted genes. The mature RNA must be exported from the nucleus to the cytoplasm. Regulation of gene expression by selection of sequences at the level of nuclear RNA might involve any or all of these stages, but the one for which we have most evidence concerns changes in splicing: some genes are expressed by means of alternative splicing patterns whose regulation controls the type of protein product (see Chapter 30).

Finally, the translation of an mRNA in the cytoplasm can be specifically controlled. There is little evidence for the employment of this mechanism in adult somatic cells, but it does occur in some embryonic situations, as described in Chapter 7. The mechanism is presumed to involve the blocking of initiation of translation of some mRNAs by specific protein factors.

But having acknowledged that control of gene expression can occur at multiple stages, and that production of RNA cannot inevitably be equated with production of protein, it is clear

that the overwhelming majority of regulatory events occur at the initiation of transcription. Regulation of tissue-specific gene transcription lies at the heart of eukaryotic differentiation; indeed, we see examples in Chapter 38 in which proteins that regulate embryonic development prove to be transcription factors. A regulatory transcription factor serves to provide common control of a large number of target genes, and we seek to answer two questions about this mode of regulation: what identifies the common target genes to the transcription factor; and how is the activity of the transcription factor itself regulated in response to intrinsic or extrinsic signals?

Response elements identify genes under common regulation

The principle that emerges from characterizing groups of genes under common control is that *they share a promoter element that is recognized by a regulatory transcription factor.* An element that causes a gene to respond to such a factor is called a response element; examples are the HSE (heat shock response element), GRE (glucocorticoid response element), SRE (serum response element).

The properties of some inducible transcription factors and the elements that they recognize are summarized in Table 29.1. Response elements have the same general characteristics as upstream elements of promoters or enhancers. They contain short consensus sequences, and copies of the response elements found in different genes are closely related, but not necessarily identical. The region bound by the factor extends for a short distance on either side of the consensus sequence. In promoters, the elements are not present at fixed distances from the startpoint, but are usually <200 bp upstream of it. The presence of a single element usually is sufficient to confer the regulatory response, but sometimes there are multiple copies.

Table 29.1 Inducible transcription factors bind to response elements that identify groups of promoters or enhancers subject to coordinate control.

Regulatory Agent	Module	Consensus	Factor
Heat shock	HSE	CNNGAANNTCCNNG	HSTF
Glucocorticoid	GRE	TGGTACAAATGTTCT	Receptor
Phorbol ester	TRE	TGACTCA	AP1
Serum	SRE	CCATATTAGG	SRF

Response elements may be located in promoters or in enhancers. Some types of elements are typically found in one rather than the other: usually an HSE is found in a promoter, while a GRE is found in an enhancer. We assume that all response elements function by the same general principle. *A gene is regulated by a sequence at the promoter or enhancer that is recognized by a specific protein. The protein functions as a transcription factor needed for RNA polymerase to initiate. Active protein is available only under conditions when the gene is to be expressed; its absence means that the promoter is not activated by this particular circuit.*

An example of a situation in which many genes are controlled by a single factor is provided by the heat shock response. This is common to a wide range of prokaryotes and eukaryotes and involves multiple controls of gene expression: an increase in temperature turns off transcription of some genes, turns on transcription of the heat shock genes, and causes changes in the translation of mRNAs. The control of the heat shock genes illustrates the differences between prokaryotic and eukaryotic modes of control. In bacteria, a new sigma factor is synthesized that directs RNA polymerase holoenzyme to recognize an alter-